

Abstract

We study minimization problems with respect to a one-parameter family of generalized relative entropies. These relative entropies, which we call relative α -entropies (denoted $\mathcal{J}_\alpha(P, Q)$), arise as redundancies under mismatched compression when cumulants of compression lengths are considered instead of expected compression lengths. These parametric relative entropies are a generalization of the usual relative entropy (Kullback-Leibler divergence). Just like relative entropy, these relative α -entropies behave like squared Euclidean distance and satisfy the Pythagorean property. We explore the geometry underlying various statistical models and its relevance to information theory and to robust statistics. The thesis consists of two parts.

In the first part, we study minimization of $\mathcal{J}_\alpha(P, Q)$ as the first argument varies over a convex set \mathbb{E} of probability distributions. We show the existence of a unique minimizer when the set \mathbb{E} is closed in an appropriate topology. We then study minimization of \mathcal{J}_α on a particular convex set, a linear family, which is one that arises from linear statistical constraints. This minimization problem generalizes the maximum Rényi or Tsallis entropy principle of statistical physics. The structure of the minimizing probability distribution naturally suggests a statistical model of power-law probability distributions, which we call an α -power-law family. Such a family is analogous to the exponential family that arises when relative entropy is minimized subject to the same linear statistical constraints.

In the second part, we study minimization of $\mathcal{J}_\alpha(P, Q)$ over the second argument. This minimization is generally on parametric families such as the exponential family or the α -power-law family, and is of interest in robust statistics ($\alpha > 1$) and in constrained compression settings ($\alpha < 1$). We show an orthogonality relationship between the α -power-law family and an associated linear family. As a consequence of this, the minimization of $\mathcal{J}_\alpha(P, \cdot)$, when the second argument

comes from an α -power-law family, can be shown to be equivalent to a minimization of $\mathcal{J}_\alpha(\cdot, R)$, for a suitable R , where the first argument comes from a linear family. The latter turns out to be a simpler problem of minimization of a quasi-convex objective function subject to linear constraints. Standard techniques are available to solve such problems, for example, via a sequence of convex feasibility problems, or via a sequence of such problems but on simpler single-constraint linear families.